PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The cost effectiveness of a structured education pulmonary rehabilitation programme for chronic obstructive pulmonary disease in primary care: The PRINCE cluster randomised trial.
AUTHORS	Gillespie, Paddy; O'Shea, Eamon; casey, dympna; Murphy, Kathy; Devane, Declan; Cooney, Adeline; Mee, Lorraine; Kirwan, Collette; McCarthy, Bernard; Newell, John

VERSION 1 - REVIEW

REVIEWER	Helen Dakin Senior Researcher Health Economics Research Centre Nuffield Department of Population Health University of Oxford UK I have no competing interests in relation to this article
REVIEW RETURNED	22-Jul-2013

THE STUDY	• Abstract: Need to state the perspective within the abstract and state the comparator as part of the objective. The abstract would be improved by stating the cost per QALY gained and cost per CRQ gained (either as well as or instead of the probability that treatment is cost-effective). • page 5, paragraph 2: it would be useful to state whether the SEPRP education sessions were done on groups of patients or 1:1 with individuals. • Page 7, lines 52-55 and page 8, lines 41-43: substantially more details are needed on the imputation. What method of imputation was conducted and why, was uncertainty around imputed values propagated into the analysis (and if so how), what software or command was used and was the imputation conducted on total costs or on the quantities of individual resources? These statements also imply that imputation of costs and QALYs was conducted separately and therefore makes no allowance for the correlations between costs and outcomes. It is also not clear whether data were imputed on CRQ and EQ-5D or just EQ-5D. It is also unclear whether values missing after patients withdrew from the study were imputed in the same way as missing questionnaires, or how deaths were taken into account in the analysis of CRQ. • It is unclear when participants completed resource use questionnaires and how the results were used to generate the quantities and costs in Table 2 and Appendix Table 3. It is unclear how resource use data were collected for patients who had died. • Page 6, line 56: it would be useful to say “evidence on resource use, EQ-5D and CRQ was collected”, rather than simply “health status”

	• Page 7, line 55-57 and page 8, lines 43-6: these statements imply that costs, CRQ and QALYs were estimated separately and therefore make no allowance for the correlations between costs and effects. If this is the case it is a major shortcoming of the analysis. It might be clearer to describe the analysis and imputation of both costs and outcomes within the same section, rather than repeating almost identical statements in the cross-section and the outcomes section. Although the distribution used is stated for costs, there is no indication of what distribution was assumed for outcomes. At a minimum, the approach used needs to be clearly described and justified. • Page 8, lines 10-23: it would be very useful to state what the range of scores is for CRQ to give some context to the magnitude of the differences observed in the trial • Page 9, line 11: The i, j and k terms in this equation and the sentence that follows should presumably be put in subscript. • Page 9, lines 29-31: It is unclear from the methods whether this analysis of incremental net benefit is part of an analysis of the joint distribution of costs and effects, or whether it was conducted separately as a univariate analysis. It is also necessary to state what distribution was used. It is also not entirely clear which of the results are informed by this analysis (presumably just the probability that treatment is cost-effective?) • Table 1: full references are needed for all sources of unit costs. The majority of drugs included in this table appear to be given as brand names; in the case of combination inhalers this may be appropriate, although it would be useful to also state the generic names. It would also be useful to state whether "prednisone" is given orally or inhaled. • Appendix table 1: is this a duplicate of information published in an earlier paper? If so, this should be clearly stated and permission is likely to be needed to replicate the table. • Appendix Table 2: it would be helpful to separate out patient expenses from healthcare service expenses in all sections of this table. At present, it is difficult to follow the calculations: particularly when the figure of €564 appears in two places.
RESULTS & CONCLUSIONS	• The results (most notably Table 2 and Appendix Table 3) are currently not clearly presented. • Table 2 and Appendix Table 3 appear to show a concerningly large difference in baseline CRQ and baseline EQ-5D utility, with the intervention group having substantially higher baseline quality of life than the control group. Given the magnitude of these differences, it is highly likely that the analysis would have been sensitive to the methods used to adjust for baseline. It is unclear which results have been adjusted for baseline. Assuming that the results in table 2 include no adjustment for baseline while the results in table 3 are adjusted suggests that adjusting for baseline using GEE eliminates the difference in QALYs, but not the difference in CRQ, which is surprising. • The study includes no sensitivity analyses. It is generally recommended that all important assumptions and methodological approaches are varied in sensitivity analysis. In particular, in this case it would have been particularly useful to explore the impact of different approaches to missing data or baseline imbalance, as well as varying the number of patients per practice and the number of practices by which the intervention costs are divided. • As with most economic evaluations, the conclusions of the analysis hinge upon the ceiling ratio used to interpret the results. The authors avoid specifying a ceiling ratio a priori since the ceiling ratio is "unknown for Ireland". However, the authors nonetheless draw

	conclusions that are implicitly based on strong judgements about the ceiling ratio: for example in the abstract they state that “strongly favourable results exist when health status was measured using the disease-specific CRQ instrument”, as well as similar statements in the discussion. It is unclear what ceiling ratio these statements are based on, although it would appear that the authors only consider ceiling ratios in excess of €5000 per one-point increase in CRQ. This ceiling ratio seems extremely high: based on the UK ceiling ratio of £20,000/QALY gained, a ceiling ratio of €5000 per one-point increase in CRQ would suggest that a 1 point increase in CRQ is equivalent to around 0.167 QALYs, which seems unlikely. It is therefore very unclear whether the conclusions are justified. The range of ceiling ratios presented in the abstract, Table 3, page 10, lines 30-35 is likely to be too high for CRQ and it may be better to present a different range of ceiling ratios for each of the two outcomes and present the CEACs on different graphs. It may also not be necessary to present the probability of being cost-effective at five ceiling ratios in four different places in the paper. The shape of the CEAC for CRQ could also be quite different from that shown in Figure 1 if additional points between 0 and €5000 were evaluated. • Table 2 and Appendix Table 3: It is not entirely clear exactly what results are presented in these table or how they were estimated. It would be useful to explicitly state whether the resource use and cost data presented in Table 2 represents the total used over the 22-week study period (rather than any other time period). It is unclear why physiotherapy, public health nurse, dietician, home help and social worker visits are recorded as percentages, rather than quantities used over the 22-week study period. Presenting such figures as percentages raises the question of how the total cost over the 22-week period was estimated from the percentage at baseline and percentage at 22 weeks: for example, does the analysis assume that anyone who stopped having physiotherapy between baseline and 22 weeks stopped halfway through and what assumptions are made about how frequently patients saw a physiotherapist? It is also unclear why Appendix table 3 is labelled “baseline (26 weeks)”: does it represent the resource use on the day of randomisation, resource use 26 weeks before the study intervention, resource use 26 weeks after the start of study intervention or the resource use that patients reported at baseline when recalling their resource use over the past 26 weeks? How are the baseline resource use quantities taken into account in the analysis? It is also unclear how the standard deviations were calculated (particularly if multiple imputation was conducted) and how (or whether) the data presented in this table include imputation of missing data. It is also unclear whether these figures were estimated in the GEE models that adjusted for baseline differences, or whether they are unadjusted figures. • Table 2 last line and page 9, line 3: it would be clearer to describe the QALYs as “unadjusted QALYs”, rather than “QALYs gained” if this figure represents the total unadjusted QALYs over the 22-week period. • Page 9, line 56: it is unclear which time point these CRQ scores represent: are they based on the measurements at 22 weeks, the average between baseline and 22 weeks or something else? • It would be useful to state the cost-effectiveness ratios from a healthcare perspective as well as from a patient+ healthcare perspective (either in Table 3 or in the text) • It would be useful to comment on the statistical significance of the adjusted differences in QALYs, CRQ, total costs, healthcare service costs and patient costs between treatment arms and give p-values. • Page 10, lines 48-51: the double negative in this sentence makes it

	very difficult to understand and possibly inaccurate. • Page 11, lines 21-31 and page 12 lines 11-13: I would anticipate that a great many studies have looked at the sensitivity of EQ-5D in moderate COPD and it would be useful to cite these here. In particular, if there is evidence finding EQ-5D to be sensitive and responsive in COPD, it would be useful to discuss this in relation to the results of the current study. • Page 11, 31-35: additional justification may be needed to conclude that other studies should use disease-specific health status measures to assess cost-effectiveness and resource allocation. This study also highlights the difficulty of identifying an appropriate ceiling ratio and drawing conclusions based on ICERs using non-preference-based measures. • The study used 2009 prices. It would be useful if the paper discussed whether (or how) prices have changed and whether the conclusions are still valid in 2013 • The 22-week time horizon is a big weakness (as the authors acknowledge). It would be useful to have additional discussion on what the CUA results might have been if results were extrapolated beyond the 22-week time horizon and on whether there is any evidence on how long the effect of interventions such as this tends to last. • Page 11, lines 55-8: the wording of this sentence implies that further follow-up is planned, although if it is the case, it is unclear why this follow-up hasn't already been finished if the main trial finished in 2011.
REPORTING & ETHICS	The following points on the EVEREST checklist have not been followed: o page 6, lines 53-54: perspective isn't justified. The perspective used in the analysis ("the perspective of the health care provider and the patient") isn't a commonly used one and it is particularly useful to justify why this perspective was used. This sentence could also be made clearer and it may be better to have one sentence stating the perspective and another sentence stating the units of health outcomes. o Page 8, lines 26-40: it is important to state which UK EQ-5D tariff was used (presumably the time trade-off tariff) and give a reference. This paragraph should also state at what time points EQ-5D was measured and briefly describe what assumptions were used to interpolate between the measurements (e.g. assuming linear changes) and between a patient's last measurement and death. It would also be useful to explicitly state that EQ-5D-3L was used. o it would be useful to explicitly describe the study as a cost utility analysis and cost-effectiveness analysis within the text o The paper should explicitly state that costs and QALYs were not discounted o No sensitivity analysis was conducted. Allowance for sampling variation does not constitute sensitivity analysis Looking at the latest checklist (CHEERS), the study also doesn't fulfil the following points: o [Explain] why the single study was a sufficient source of clinical effectiveness data o Discussion of generalisability and equity

REVIEWER	Dr Julia Walters Senior Research Fellow in Primary Health Care UNIVERSITY OF TASMANIA
REVIEW RETURNED	02-Aug-2013

GENERAL COMMENTS	This interesting study has added useful knowledge to the important topic of the effectiveness of pulmonary rehabilitation in COPD, different formats, effects and cost effectiveness. It is well conducted and reported and my comments relate mainly to optimizing understanding for the clinician reader. The CHEERS Checklist of items to include when reporting economic evaluations of health interventions is endorsed by BMJ Open, yet the paper is accompanied by the EVEREST statement checklist for a Health economics paper instead, which was a little confusing. On balance I prefer the CHEERS checklist although relevant items in EVEREST were completed. **Research question clear definition, study design** The paper reports data on cost effectiveness from a pulmonary rehabilitation cluster RCT; this is referenced as ref 9, which is described as "forthcoming". Does this mean it is accepted for publication or is still under review? Could you clarify please. [9. Murphy K, Casey D, Devane D, et al. The effectiveness of a structured education pulmonary rehabilitation programme for improving the health status of people with moderate and severe chronic obstructive pulmonary disease in primary care: The PRINCE cluster randomised trial.2013.*Thorax*(*forthcoming*)] The introduction contained details of the PRINCE RCT that belong better in Methods section. I suggest moving this description to Methods under a subheading "details of PRINCE cluster RCT, i.e. from page 5 line 20 starting "Full details of the study methods…." To Page 6 line 11, ending "Notably however, concerns arose as the confidence intervals did not exclude differences in effect that were pre-specified as clinically insignificant.[9]" The intervention when described in the main paper and the abstract is described as a "structured education pulmonary rehabilitation programme (SEPRP)" whereas the reader needs to understand its components relative to "traditional" pulmonary rehabilitation, against which it is being considered as an alternative. Thus it needs to be clear if it is a structured education and supervised exercise pulmonary rehabilitation programme, or what was the form of the exercise component. The opening sentence in paragraph 2 page 5 should then contain a brief and accurate definition of the intervention, (line 18) such as, "The PRINCE study sought to examine the clinical and cost effectiveness of a *structured education and supervised exercise pulmonary rehabilitation programme (SEPRP)* for COPD delivered at

	the level of general practice in Ireland." The closing sentence (Page 6 lines 38-43) of the Introduction should then define the outcome measure used, the Chronic Respiratory Questionnaire (CRQ) **Description of methods** I found it difficult to understand some details of the methods used in the cluster RCT that are relevant to understanding the cost-effectiveness analysis, but if the paper on the main results is published before these results, this will help the reader. The methods used in the cost-effectiveness analysis reported here are generally well described. Authors state (on page 7 lines 52-55), that they used imputation to estimate healthcare costs and patient costs in the case of missing values for some individuals at follow up. Could you clarify how many participants had missing data and whether any sensitivity analyses were done with varying estimates? In "Effectiveness analysis", page 8 lines 10-23, a description of the CRQ is given and referenced [Ref 10]. Could you clarify how the three CRQ aggregate scores have been calculated, as this is not included in the original measure as referenced. Similarly to the above comments, on page 8 lines 41-46, could you clarify how many participants had missing data and whether any sensitivity analyses were done with varying estimates? In the Sub-Section "Overview", in the Methods section, authors state the use of a time horizon of 22 weeks. Could you explain why this period was used and why it was appropriate. (Line 53 page 6) If using perspectives of both the healthcare provider and the patient, should "perspectives" not be plural? (line 58, page 6) I was puzzled initially by "practice note searches" and assume you mean practice records? Were these paper or electronic records? **Results:** These are clearly described, although placing the Table of "Characteristics of clusters (general practices) and baseline demographic and clinical characteristics of COPD patients" as an appendix seems a pity. I would prefer to have that data in the main paper. It is also referred to as "Appendix Table 3" in the results section (page 9 line 46), whereas there is a Table 3 on page 17 and the Table in the appendix is labeled "Appendix Table 1".

	Table 3 (page 17) add “per patient” to cost figures for healthcare resources and patient resources. **Discussion** The difference in incremental cost effectiveness ratios based on the CRQ and QALY gained from the health utility measure highlighted by the authors are very interesting. Although they discuss some explanations and compare the results with other studies, I think this section could be expanded; especially to highlight the difference in target population for the SEPRP programme (more moderate COPD than hospital recruited programmes even if actually delivered in a non-hospital setting). It would be useful to give more details on studies reporting reduced hospital utilization (mostly in severe, very severe COPD) and compare directly with any other ICER estimates. Discussion of cost effectiveness of other rehabilitation/self management programmes delivered in primary care in diabetes for example DESMOND (Gillett et al BMJ 2010). **Abstract** In the objective here the intervention should be clarified, again using content description of the intervention, whether it was a structured education and supervised exercise pulmonary rehabilitation programme. Participants: useful to know here 69% with moderate COPD Intervention- use an accurate description Main outcome measure: Include the basis for QALYs, ie the generic EQ5D. Results: clarify the mean increases in healthcare and patient costs are **per patient.** Conclusions: Use an accurate description of the intervention. If you have defined use of the generic EQ5D in the outcome measures this can be deleted here.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1's Comments

R1 Point 1: Abstract: Need to state the perspective within the abstract and state the comparator as part of the objective. The abstract would be improved by stating the cost per QALY gained and cost per CRQ gained (either as well as or instead of the probability that treatment is cost-effective).

RESPONSE: The abstract has now been updated. The perspective of the cost effectiveness analysis is now explicitly that of the health care provider.

R1 Point 2: page 5, paragraph 2: it would be useful to state whether the SEPRP education sessions were done on groups of patients or 1:1 with individuals.

RESPONSE: The SEPRP sessions were group-based. This has been updated throughout the paper.

R1 Point 3: Page 7, lines 52-55 and page 8, lines 41-43: substantially more details are needed on the imputation. What method of imputation was conducted and why, was uncertainty around imputed values propagated into the analysis (and if so how), what software or command was used and was the imputation conducted on total costs or on the quantities of individual resources? These statements also imply that imputation of costs and QALYs was conducted separately and therefore makes no allowance for the correlations between costs and outcomes. It is also not clear whether data were imputed on CRQ and EQ-5D or just EQ-5D. It is also unclear whether values missing after patients withdrew from the study were imputed in the same way as missing questionnaires, or how deaths were taken into account in the analysis of CRQ.

RESPONSE: We now include data on the level of missing data in the footnotes for Table 2, we now include the following text:

"Note: Eight patients (6 intervention and 2 control) who died over the course of the study were excluded from the analysis. Completeness of cost data: Intervention - 99% for on primary care utilisation, 99% for secondary care utilisation, 80% for community care utilisation, 99% for medication utilisation, 80% for oxygen therapy utilisation, and 78% for Total Healthcare Cost. Control: 97% 97%, 78%, 97%, 78% and 78% respectively.. Completeness of effect data: Intervention - 80% for CRQ, 80% for EQ5D and 80% for QALY scores. Control - 78%, 78% and 78% (N=134) respectively."

Given the limited length of follow up, we did not analyse data for those 8 patients who died.

With respect to imputation, we imputed values using the UVIS command in STATA. On the cost side, we imputed individual resource use at follow up for patients alive at follow up conditional on age, gender, and treatment arm. This data was used to calculate Total Healthcare Cost at follow up, as we had a 78% response rate for both Intervention and Controls in the complete case analysis. On the effect side, we imputed CRQ and EQ5D at follow up for patients alive at follow up (note: we had complete data for 80% for intervention and 78% for control patients) using the UVIS code in STATA. As stated, in the paper, this was imputed, conditional on age, gender, and treatment arm. With respect to the EQ5D, These estimates were combined with the baseline EQ5D values to estimate, using the area under the curve method, QALYs at 22 weeks.

The reviewer is correct in stating that resource use and EQ5D values were imputed separately and therefore make no allowance for the correlations between costs and outcomes. Importantly, we decided to impute the limited amount of missing data in this way, and to account for the correlation between cost and outcome data in the incremental cost effectiveness analysis which follows. Indeed, this was a pragmatic decision, given that we impute value for the long list of resource use categories.

Nonetheless, and given the relatively low level of missing data, we believe this approach to be appropriate. Furthermore, in the updated paper, we present results for the complete cases only analysis.

R1 Point 4: It is unclear when participants completed resource use questionnaires and how the results were used to generate the quantities and costs in Table 2 and Appendix Table 3. It is unclear how resource use data were collected for patients who had died.

RESPONSE: We we now include further information on this process in the cost analysis. As stated above, we excluded patients who died from the final analysis.

R1 Point 5: Page 6, line 56: it would be useful to say “evidence on resource use, EQ-5D and CRQ was collected”, rather than simply “health status”

RESPONSE: This has now been updated.

R1 Point 6: Page 7, line 55-57 and page 8, lines 43-6: these statements imply that costs, CRQ and QALYs were estimated separately and therefore make no allowance for the correlations between costs and effects. If this is the case it is a major shortcoming of the analysis. It might be clearer to describe the analysis and imputation of both costs and outcomes within the same section, rather than repeating almost identical statements in the cross-section and the outcomes section.

RESPONSE: Given the data, now included in the paper with respect to the missing data, we don’t believe this to be the case. Furthermore, the cost effectiveness results, which are the main findings from the study, are based on the incremental net benefit analysis. The estimation of these results do explicitly take into account the correlation between the costs and effects which, as the reviewer states, is of primary importance in the analysis of cost effectiveness. Therefore, we do address this important issue in our analysis.

For presentation purposes, we felt it best to present the methods and results for the cost analysis and effectiveness analysis separately. We believe this approach gives the reader a clearer picture of how we estimate the individual components for costs and effects, before we combined these two variables to undertake the cost effectiveness analysis.

R1 Point 7: Although the distribution used is stated for costs, there is no indication of what distribution was assumed for outcomes. At a minimum, the approach used needs to be clearly described and justified.

RESPONSE: The regression models for the incremental effectiveness (CRQ and QALYs) analysis and the incremental cost effectiveness analysis (Net Benefit) were estimated using GEE regression models, which assumed Gaussian variance functions. This is now made explicit in the paper.
The Gaussian based models were chosen on the basis of the QIC criterion.

R1 Point 8: Page 8, lines 10-23: it would be very useful to state what the range of scores is for CRQ to give some context to the magnitude of the differences observed in the trial

RESPONSE: For consistency, we present means and standard deviations for descriptive statistics. We did attempt to include data on the range of all variables in the updated version of Table 2; however, this did not read well. Further information on the range of values for CRQ is presented in the clinical paper by Casey et al (2013). Moreover, the following may be inserted if preferable to the editorial board:

“In terms of disease-specific health status, mean CRQ Total score per patient at follow up was 19.10 (SD: 4.83, range 7.7 to 28.4)) in the control arm and 20.82 (SD: 3.88, range 8.6 to 27.5) in the intervention arm.”

R1 Point 9: Page 9, line 11: The i, j and k terms in this equation and the sentence that follows should presumably be put in subscript.

RESPONSE: This has now been updated

R1 Point 10: Page 9, lines 29-31: It is unclear from the methods whether this analysis of incremental net benefit is part of an analysis of the joint distribution of costs and effects, or whether it was conducted separately as a univariate analysis. It is also necessary to state what distribution was used. It is also not entirely clear which of the results are informed by this analysis (presumably just the probability that treatment is cost-effective?)

RESPONSE: The paper now hopefully clarifies this issue. As stated in the Overview section of the methods: “… we adopt statistical techniques which recognise both the clustering and correlation of cost and effect data”, which is of central importance in the analysis of cost and effect data collected alongside cluster RCTs. This is now explicitly stated in the “Cost Effectiveness Analysis” section of the Methods

To clarify, the cost effectiveness results, which are the main findings from the study, are based on the incremental net benefit analysis. The estimation of the net benefit statistic allows us to explicitly take into account the correlation between the costs and effects which is of primary importance in the analysis of cost effectiveness. Furthermore, the estimation of incremental net benefit statistics using multilevel GEE models allow us to take account of the clustering in the dataset. Finally, it is the results from these regression models that are directly used to estimate the probability values which are presented in the CEACs.

As stated above, we do now also state in the paper that the regression models for the cost effectiveness analysis (Net Benefit) were estimated using GEE regression models, assuming a Gaussian variance functions. The Gaussian based models were chosen on the basis of their QIC measures.

R1 Point 11: Table 1: full references are needed for all sources of unit costs. The majority of drugs included in this table appear to be given as brand names; in the case of combination inhalers this may be appropriate, although it would be useful to also state the generic names. It would also be useful to state whether “prednisone” is given orally or inhaled.

RESPONSE: This was purely a space-related issue. We simply don’t have sufficient room in the table to include full information on all data sources. We do present further information on these sources in the footnotes to the table. We had extensive data on the drug prescriptions. We now include the full list of Generic and Brand names. For each drug, we identified unit costs from the Monthly Index of Medical Specialities (MIMS) for Ireland. Prednisone was orally administered.

R1 Point 12: Appendix table 1: is this a duplicate of information published in an earlier paper? If so, this should be clearly stated and permission is likely to be needed to replicate the table.

RESPONSE: This table is applicable to both our clinical and economic papers. We now reference the original paper for this table.

R1 Point 13: Appendix Table 2: it would be helpful to separate out patient expenses from healthcare

service expenses in all sections of this table. At present, it is difficult to follow the calculations: particularly when the figure of €564 appears in two places.

RESPONSE: This is a valid point and we now have updated the Table.

R1 Point 14: The results (most notably Table 2 and Appendix Table 3) are currently not clearly presented.

RESPONSE: Both tables have been altered significantly and are, hopefully, now clearly presented. Notable, these tables included the raw data estimates for the variables of interest. The 'Totals' are now presented in the new version of Table 3.

R1 Point 15: Table 2 and Appendix Table 3 appear to show a concerningly large difference in baseline CRQ and baseline EQ-5D utility, with the intervention group having substantially higher baseline quality of life than the control group. Given the magnitude of these differences, it is highly likely that the analysis would have been sensitive to the methods used to adjust for baseline. It is unclear which results have been adjusted for baseline. Assuming that the results in table 2 include no adjustment for baseline while the results in table 3 are adjusted suggests that adjusting for baseline using GEE eliminates the difference in QALYs, but not the difference in CRQ, which is surprising.

RESPONSE: As stated, Table 2 and Appendix Table 3 has been updated and present the raw unadjusted results for these variables. The results in Table 3 present the results from the regression analyses which do adjust for the baseline values and the variation in these values (in CRQ, EQ5D and Total Cost). We present the results in this manner to ensure the reader has a clear indication of the methods adopted and the results obtained. Notably, given the significant improvements detected in CRQ, unlike those for QALYS gained, we believe this result to be consistent.

R1 Point 16: The study includes no sensitivity analyses. It is generally recommended that all important assumptions and methodological approaches are varied in sensitivity analysis. In particular, in this case it would have been particularly useful to explore the impact of different approaches to missing data or baseline imbalance, as well as varying the number of patients per practice and the number of practices by which the intervention costs are divided.

RESPONSE: This is a valid point. Space considerations were the main reason for our excluding an explicit sensitivity analysis. We now present the follow sensitivity analyses in the appendix:
(1) Complete Case Analysis
(2) Intervention Costs – More Patients Versus Less Patients

R1 Point 17: As with most economic evaluations, the conclusions of the analysis hinge upon the ceiling ratio used to interpret the results. The authors avoid specifying a ceiling ratio a priori since the ceiling ratio is "unknown for Ireland". However, the authors nonetheless draw conclusions that are implicitly based on strong judgements about the ceiling ratio: for example in the abstract they state that "strongly favourable results exist when health status was measured using the disease-specific CRQ instrument", as well as similar statements in the discussion. It is unclear what ceiling ratio these statements are based on, although it would appear that the authors only consider ceiling ratios in excess of €5000 per one-point increase in CRQ. This ceiling ratio seems extremely high: based on the UK ceiling ratio of £20,000/QALY gained, a ceiling ratio of €5000 per one-point increase in CRQ would suggest that a 1 point increase in CRQ is equivalent to around 0.167 QALYs, which seems unlikely. It is therefore very unclear whether the conclusions are justified. The range of ceiling ratios presented in the abstract, Table 3, page 10, lines 30-35 is likely to be too high for CRQ and it may be better to present a different range of ceiling ratios for each of the two outcomes and present the

CEACs on different graphs. It may also not be necessary to present the probability of being cost-effective at five ceiling ratios in four different places in the paper. The shape of the CEAC for CRQ could also be quite different from that shown in Figure 1 if additional points between 0 and €5000 were evaluated.

RESPONSE: This is a fair point and we now tone down the statement with respect to 'strong' evidence of cost effectiveness with respect to the CRQ results. Nonetheless, the ceiling ratios per QALY gained presented provide a useful range for comparison, given the lack of an implicit or explicit value for Ireland, and the current evidence base with respect to this type of health economic literature for Ireland. We do take the reviewer's point in relation to our applying the same ceiling rates per unit increase in CRQ gained, and we now explicitly mention this point in the paper.

" .. Notably, the ceiling ratios per QALY gained presented provide a useful range for comparison, given the lack of an implicit or explicit value for Ireland, and the current evidence base with respect to this type of health economic literature for Ireland. Furthermore, the approach to applying the same ceiling rates per unit increase in CRQ gained is problematic as these values may or may not be much lower than those presented"

Having said that, we have no evidence on what this range of ceiling ratios should be it seems just as arbitrary to adopt a range of values for the ceiling ratio per CRQ unit. Indeed, whilst taking this point as valid, it should not be lost that there were significant improvements in CRQ, and it is these which derive the positive results in this case.

R1 Point 18: Table 2 and Appendix Table 3: It is not entirely clear exactly what results are presented in these table or how they were estimated. It would be useful to explicitly state whether the resource use and cost data presented in Table 2 represents the total used over the 22-week study period (rather than any other time period).

RESPONSE: These tables have been updated and include raw, unadjusted data.

R1 Point 19: It is unclear why physiotherapy, public health nurse, dietician, home help and social worker visits are recorded as percentages, rather than quantities used over the 22-week study period. Presenting such figures as percentages raises the question of how the total cost over the 22-week period was estimated from the percentage at baseline and percentage at 22 weeks: for example, does the analysis assume that anyone who stopped having physiotherapy between baseline and 22 weeks stopped halfway through and what assumptions are made about how frequently patients saw a physiotherapist?

RESPONSE: This has now been updated to include the mean (and standard deviation) for these resource categories.

R1 Point 20: It is also unclear why Appendix table 3 is labelled "baseline (26 weeks)": does it represent the resource use on the day of randomisation, resource use 26 weeks before the study intervention, resource use 26 weeks after the start of study intervention or the resource use that patients reported at baseline when recalling their resource use over the past 26 weeks?

RESPONSE: This reports resource use at randomisation for the prior 26 weeks. This stated in the Methods section and now is included in the Appendix.

How are the baseline resource use quantities taken into account in the analysis? It is also unclear how the standard deviations were calculated (particularly if multiple imputation was conducted) and how (or whether) the data presented in this table include imputation of missing data. It is also unclear

whether these figures were estimated in the GEE models that adjusted for baseline differences, or whether they are unadjusted figures.

RESPONSE: The resource use data in Appendix Table 3 are raw, unadjusted estimates. The baseline resource use data presented in Appendix Table 3 were used to estimate the Baseline Total Cost Variable. This variable was included as a covariate in the GEE regression models for Total Cost at follow up and Net Benefit at Follow up. Similarly, the baseline CRQ and baseline EQ5D estimates were included as covariate in the GEE regression models at Follow Up. This is stated in the paper in the methods section.

R1 Point 21: Table 2 last line and page 9, line 3: it would be clearer to describe the QALYs as "unadjusted QALYs", rather than "QALYs gained" if this figure represents the total unadjusted QALYs over the 22-week period.

RESPONSE: This has now been addressed. We explicitly refer to this as 'raw' or unadjusted data

R1 Point 22: Page 9, line 56: it is unclear which time point these CRQ scores represent: are they based on the measurements at 22 weeks, the average between baseline and 22 weeks or something else?

RESPONSE: This is now clarified by the line in the text.

"In terms of disease-specific health status, mean CRQ Total score per patient at follow up…"

R1 Point 23: It would be useful to state the cost-effectiveness ratios from a healthcare perspective as well as from a patient+ healthcare perspective (either in Table 3 or in the text)

RESPONSE: We limited our analysis now to the health care provider perspective.

R1 Point 24: It would be useful to comment on the statistical significance of the adjusted differences in QALYs, CRQ, total costs, healthcare service costs and patient costs between treatment arms and give p-values.

RESPONSE: We now add the p-values as suggested.

R1 Point 25: Page 10, lines 48-51: the double negative in this sentence makes it very difficult to understand and possibly inaccurate.

RESPONSE: We have changed this sentence.

"Moreover, the confidence intervals for thedisease-specific analysis included differences in effect that were deemed clinically insignificant.[9]"

R1 Point 26: Page 11, lines 21-31 and page 12 lines 11-13: I would anticipate that a great many studies have looked at the sensitivity of EQ-5D in moderate COPD and it would be useful to cite these here. In particular, if there is evidence finding EQ-5D to be sensitive and responsive in COPD, it would be useful to discuss this in relation to the results of the current study.

RESPONSE: We now refer to such studies, and what our paper adds, in the Discussion.

"Indeed, our findings can be added to those of existing studies which explore how the adoption of generic rather than disease-specific measures in this context may lead to the underestimation of

treatment benefits, biased cost effectiveness results, and ill-informed policy decisions.[12, 13]”

R1 Point 27: Page 11, 31-35: additional justification may be needed to conclude that other studies should use disease-specific health status measures to assess cost-effectiveness and resource allocation. This study also highlights the difficulty of identifying an appropriate ceiling ratio and drawing conclusions based on ICERs using non-preference-based measures.

RESPONSE: This is a good point. We now include this line in the discussion:

“Moreover, this study further highlights the difficulty of identifying an appropriate ceiling ratio and drawing conclusions based on ICERs using non-preference-based measures”

R1 Point 28: The study used 2009 prices. It would be useful if the paper discussed whether (or how) prices have changed and whether the conclusions are still valid in 2013

RESPONSE: This is now addressed in the Discussion.

“It should also be noted that the use of 2009 prices in the analysis may have inflated costs. Medical inflation has fallen in the period since then, which would also likely contribute to an improvement in the cost effectiveness results into the future.”

R1 Point 29: The 22-week time horizon is a big weakness (as the authors acknowledge). It would be useful to have additional discussion on what the CUA results might have been if results were extrapolated beyond the 22-week time horizon and on whether there is any evidence on how long the effect of interventions such as this tends to last.

RESPONSE: This is now addressed in more detail in the Discussion.

“The short-time horizon for our study is, therefore, a significant weakness to exploring the sustainability of the intervention. Extending the time horizon would likely improve the cost-effectiveness of the intervention, linked to lower hospital admissions, if the evidence of other studies can be used as a guide to future resource use in Ireland. ”

R1 Point 30: Page 11, lines 55-8: the wording of this sentence implies that further follow-up is planned, although if it is the case, it is unclear why this follow-up hasn’t already been finished if the main trial finished in 2011.

RESPONSE: This process is not currently planned.

The following points on the EVEREST checklist have not been followed:

R1 Point 31: page 6, lines 53-54: perspective isn’t justified. The perspective used in the analysis (“the perspective of the health care provider and the patient”) isn’t a commonly used one and it is particularly useful to justify why this perspective was used. This sentence could also be made clearer and it may be better to have one sentence stating the perspective and another sentence stating the units of health outcomes.

RESPONSE: We now explicitly state that we employ a health care provider perspective.

R1 Point 32: Page 8, lines 26-40: it is important to state which UK EQ-5D tariff was used (presumably

the time trade-off tariff) and give a reference. This paragraph should also state at what time points EQ-5D was measured and briefly describe what assumptions were used to interpolate between the measurements (e.g. assuming linear changes) and between a patient's last measurement and death. It would also be useful to explicitly state that EQ-5D-3L was used.

RESPONSE: We have now added the relevant information and references.

R1 Point 33: it would be useful to explicitly describe the study as a cost utility analysis and cost-effectiveness analysis within the text

RESPONSE: We now do so.

R1 Point 34: The paper should explicitly state that costs and QALYs were not discounted

RESPONSE: We now do so.

R1 Point 35: No sensitivity analysis was conducted. Allowance for sampling variation does not constitute sensitivity analysis

RESPONSE: We now include a sensitivity analysis, as outlined above.

R1 Point 36: Looking at the latest checklist (CHEERS), the study also doesn't fulfil the following points:

R1 Point 37: [Explain] why the single study was a sufficient source of clinical effectiveness data

RESPONSE: The design of the clinical study guided the economic study.

R1 Point 38: Discussion of generalisability and equity

RESPONSE: We do not explicitly assess equity or generalizability in the paper. From an equity perspective, the intervention was made available to all patients, regardless of socioeconomic background. The study was based on a well-designed randomised controlled trial. Issues with respect to internal and external validity are discussed in the clinical paper by Casey et al (2013).

Reviewer 2's Comments

R2 Point 1:The paper reports data on cost effectiveness from a pulmonary rehabilitation cluster RCT; this is referenced as ref 9, which is described as "forthcoming". Does this mean it is accepted for publication or is still under review? Could you clarify please. [9. Murphy K, Casey D, Devane D, et al. The effectiveness of a structured education pulmonary rehabilitation programme for improving the health status of people with moderate and severe chronic obstructive pulmonary disease in primary care: The PRINCE cluster randomised trial.2013.Thorax(forthcoming)]

RESPONSE: Thisstudy has now been published. The revised version of the paper now includes the updated reference.

Casey D, Murphy K, Devane D, Cooney A, Mee L, McCarthy B, Newell J, Scarrott C, O' Shea E, Gillespie P, Kirwan C, Murphy A. The effectiveness of a structured education pulmonary rehabilitation programme for improving the health status of people with moderate and severe chronic obstructive pulmonary disease in primary care: The PRINCE cluster randomised trial. Thorax, 2013; DOI: 10.1136/thoraxjnl-2012-203103

R2 Point 2: The introduction contained details of the PRINCE RCT that belong better in Methods section. I suggest moving this description to Methods under a subheading “details of PRINCE cluster RCT, i.e. from page 5 line 20 starting “Full details of the study methods….” To Page 6 line 11, ending “Notably however, concerns arose as the confidence intervals did not exclude differences in effect that were pre-specified as clinically insignificant.[9]”

RESPONSE: We have moved this section, as requested, to the Methods.

R2 Point 3: The intervention when described in the main paper and the abstract is described as a “structured education pulmonary rehabilitation programme (SEPRP)” whereas the reader needs to understand its components relative to “traditional” pulmonary rehabilitation, against which it is being considered as an alternative. Thus it needs to be clear if it is a structured education and supervised exercise pulmonary rehabilitation programme, or what was the form of the exercise component.

RESPONSE: The comparators in this study are the intervention, which consisted of SEPRP programme, and the control which consisted of usual practice in primary care. While pulmonary rehabilitation is offered in the hospital setting, there is no such equivalent in primary care. This study compares the SEPRP intervention to existing usual care for COPD in the community setting. To this end, the wide range of data we present with respect to the use of primary care, secondary care, community care, and medications identifies, in actual practice, usual care consists off in this setting in Ireland. .

R2 Point 4:The opening sentence in paragraph 2 page 5 should then contain a brief and accurate definition of the intervention, (line 18) such as, “The PRINCE study sought to examine the clinical and cost effectiveness of a structured education and supervised exercise pulmonary rehabilitation programme (SEPRP) for COPD delivered at the level of general practice in Ireland.”
The closing sentence (Page 6 lines 38-43) of the Introduction should then define the outcome measure used, the Chronic Respiratory Questionnaire (CRQ)

RESPONSE: We have incorporated these changes in the updated version.

R2 Point 5:I found it difficult to understand some details of the methods used in the cluster RCT that are relevant to understanding the cost-effectiveness analysis, but if the paper on the main results is published before these results, this will help the reader.

RESPONSE: The main clinical paper is now published and we refer the reader to this study for further information.

R2 Point 6: The methods used in the cost-effectiveness analysis reported here are generally well described. Authors state (on page 7 lines 52-55), that they used imputation to estimate healthcare costs and patient costs in the case of missing values for some individuals at follow up. Could you clarify how many participants had missing data and whether any sensitivity analyses were done with varying estimates?

RESPONSE: We now include full information on missing data. In sensitivity analysis we present cost effectiveness results for the complete case analysis.

R2 Point 7: In “Effectiveness analysis”, page 8 lines 10-23, a description of the CRQ is given and referenced [Ref 10]. Could you clarify how the three CRQ aggregate scores have been calculated, as this is not included in the original measure as referenced.

RESPONSE: This is a relatively straight forward process which involves summing the scores for the individual domains. This is now included.

“Based on patient responses, three CRQ aggregate scores can be calculated: (i) CRQ Physical, which is an aggregate of the dyspnoea and fatigue domains; (ii) CRQ Psychological, which is an aggregate of the emotional function and mastery domains; and (iii) CRQ Total, which is an aggregate of all four domains.”

R2 Point 8: Similarly to the above comments, on page 8 lines 41-46, could you clarify how many participants had missing data and whether any sensitivity analyses were done with varying estimates?

RESPONSE: We now include full information on missing data. In sensitivity analysis we present cost effectiveness results for the complete case analysis.

R2 Point 9: In the Sub-Section “Overview”, in the Methods section, authors state the use of a time horizon of 22 weeks. Could you explain why this period was used and why it was appropriate. (Line 53 page 6) If using perspectives of both the healthcare provider and the patient, should “perspectives” not be plural?

RESPONSE: The time horizon of the analysis was based on the follow up period of the trial. This is short and we now acknowledge this as a limitation of the analysis. The line on the perspective of the analysis has been changed in the updated version of the paper.

R2 Point 10: (line 58, page 6) I was puzzled initially by “practice note searches” and assume you mean practice records? Were these paper or electronic records?

RESPONSE: Yes, these were electronic practice records.

R2 Point 11: Results: These are clearly described, although placing the Table of “Characteristics of clusters (general practices) and baseline demographic and clinical characteristics of COPD patients” as an appendix seems a pity. I would prefer to have that data in the main paper.

RESPONSE: As this information is presented in the clinical paper, we’ve placed it in an appendix. However, we would be happy to move it to the main paper if preferable to the editorial board.

R2 Point 12: It is also referred to as “Appendix Table 3” in the results section (page 9 line 46), whereas there is a Table 3 on page 17 and the Table in the appendix is labeled “Appendix Table 1”.

RESPONSE: This has been updated in the revised paper.

R2 Point 13: Table 3 (page 17) add “per patient” to cost figures for healthcare resources and patient resources.

RESPONSE: This has been updated in the revised paper.

R2 Point 14: The difference in incremental cost effectiveness ratios based on the CRQ and QALY gained from the health utility measure highlighted by the authors are very interesting. Although they discuss some explanations and compare the results with other studies, I think this section could be expanded; especially to highlight the difference in target population for the SEPRP programme (more moderate COPD than hospital recruited programmes even if actually delivered in a non-hospital setting).

RESPONSE: In the discussion, we now to more adequately attempt to address this issue.

R2 Point 15: It would be useful to give more details on studies reporting reduced hospital utilization (mostly in severe, very severe COPD) and compare directly with any other ICER estimates. Discussion of cost effectiveness of other rehabilitation/self management programmes delivered in primary care in diabetes for example DESMOND (Gillett et al BMJ 2010).

RESPONSE: We now make reference to this point in the Discussion in the revised paper.

R2 Point 16: In the objective here the intervention should be clarified, again using content description of the intervention, whether it was a structured education and supervised exercise pulmonary rehabilitation programme.

RESPONSE: The Abstract in now updated.

R2 Point 17: Participants: useful to know here 69% with moderate COPD

RESPONSE: The Abstract in now updated

R2 Point 18: Intervention- use an accurate description

RESPONSE: The Abstract in now updated

R2 Point 19: Main outcome measure: Include the basis for QALYs, ie the generic EQ5D.

RESPONSE: The Abstract in now updated

R2 Point 20: Results: clarify the mean increases in healthcare and patient costs are per patient.

RESPONSE: The Abstract in now updated

R2 Point 21: Conclusions: Use an accurate description of the intervention. If you have defined use of the generic EQ5D in the outcome measures this can be deleted here.

RESPONSE: The Abstract in now updated

VERSION 2 – REVIEW

REVIEWER	Dakin, Helen University of Oxford, Health Economics Research Centre, Department of Public Health
REVIEW RETURNED	25-Sep-2013

GENERAL COMMENTS	Comments to authors The key messages and conclusions of the abstract are phrased in an unclear way: e.g. “favourable cost-effectiveness results exist when health status was measured using the disease-specific CRQ instrument”, rather than a more precise statement such as “Analyses using the CRQ suggested that SEPRP was cost-effective if society is willing to pay at least €850 per one-point increase in CRP”. In particular, page 12, line 10, page 14 line

	21, page 2 lines 19 & 25 and page 4 line 9 should state what ceiling ratio is being used: e.g. by adding "if society is willing to pay at least 850 Euros per 1-unit increase in CRQ" after "health status". In practice, it is highly likely that society's willingness to pay, or the amount that the healthcare system can afford is far lower than 850 Euros per 1 point increase in CRQ, which could mean that the conclusions are the same for CRQ as for QALYs. Page 7, line 20: The paper should explicitly state what follow-up point(s) CRQ, EQ-5D and resource use are measured? (measured only at baseline and 22 weeks?) It would also be useful to explicitly state that the follow-up is at 22 weeks on page 11 (line 3) and page 10, line 45. Brief information on the methods for imputation needs to be added into page 8, lines 32-40 and page 9 lines 28-30. At a minimum, it is necessary to say that uvis was used and that utilities and each type of resource use were imputed separately, although it would also be useful to state the variables included as predictors of missing data. How many imputed datasets were generated for each variable? If more than one, how were these combined? Page 10, line 53 & page 11, lines 3 & 8: it would be useful to state that these figures are unadjusted; similarly, it would be useful to state on page 11, line 13 that these figures are adjusted for baseline utility, etc. In particular, this will help explain to readers why the incremental QALY gain (0.002) is not equal to the difference in QALYs between arms (0.337-0.305). Table 2: it would be useful to explicitly state what is meant by "raw data": i.e. to explicitly state whether the data shown here include no imputation and no adjustment for imbalance in baseline utility The study excludes patients who died, which could introduce bias (particularly as three quarters of the deaths were in the control group). The paper should justify excluding such patients and discuss what impact this could have had on conclusions. Table 2, lines 39-45: it would be useful to state whether these are at follow-up (rather than baseline), for consistency with EQ-5D and to avoid confusion. Table 3: footnotes should be added to make it clear which of the values shown in results include imputed values and which include adjustment for which covariates. For example, do all results shown in this table include imputed values, while all differences and all of

	the results below the heading of "COST EFFECTIVENESS ANALYSIS" include adjustment for baseline CRQ or EQ5D score, baseline, healthcare cost and clustering? It would also be very useful to state that the cost-effectiveness analysis results are based only on the healthcare costs (not the patient costs). Table 2 and Appendix Table 3: it would be useful to expand these table headers or add a footnote to make it 100% clear what data are being presented: for example, in Appendix table 3, it would be much clearer to say "raw data estimates at baseline (including costs in the 26 weeks leading up to randomisation)" rather than "raw data estimates at baseline (26 weeks pre-randomisation)", since the latter incorrectly implies that patients were surveyed 26 weeks before they were randomised. Similarly, in Table 2, it would be clearer to have a footnote that explicitly stated that resource use captures the period from randomisation to week 22.

VERSION 2 – AUTHOR RESPONSE

Reviewer Comment 1:

The key messages and conclusions of the abstract are phrased in an unclear way: e.g. “favourable cost-effectiveness results exist when health status was measured using the disease-specific CRQ instrument”, rather than a more precise statement such as “Analyses using the CRQ suggested that SEPRP was cost-effective if society is willing to pay at least €850 per one-point increase in CRP”. In particular, page 12, line 10, page 14 line 21, page 2 lines 19 & 25 and page 4 line 9 should state what ceiling ratio is being used: e.g. by adding "if society is willing to pay at least 850 Euros per 1-unit increase in CRQ" after "health status". In practice, it is highly likely that society’s willingness to pay, or the amount that the healthcare system can afford is far lower than 850 Euros per 1 point increase in CRQ, which could mean that the conclusions are the same for CRQ as for QALYs.

Response: We now adopt the suggested statement at each point in the paper.

Reviewer Comment 2:

Page 7, line 20: The paper should explicitly state what follow-up point(s) CRQ, EQ-5D and resource use are measured? (measured only at baseline and 22 weeks?) It would also be useful to explicitly state that the follow-up is at 22 weeks on page 11 (line 3) and page 10, line 45.

Response: This has now been updated.

Reviewer Comment 3:

Brief information on the methods for imputation needs to be added into page 8, lines 32-40 and page 9 lines 28-30. At a minimum, it is necessary to say that uvis was used and that utilities and each type of resource use were imputed separately, although it would also be useful to state the variables included as predictors of missing data. How many imputed datasets were generated for each variable? If more than one, how were these combined?

Response: We include reference to the uvis command, process, and the predictors in the paper. We

acknowledge the limitation of imputing costs and effects separately in the discussion.

Reviewer Comment 4:

Page 10, line 53 & page 11, lines 3 & 8: it would be useful to state that these figures are unadjusted; similarly, it would be useful to state on page 11, line 13 that these figures are adjusted for baseline utility, etc. In particular, this will help explain to readers why the incremental QALY gain (0.002) is not equal to the difference in QALYs between arms (0.337-0.305).

Response: This has now been updated.

Reviewer Comment 5:

Table 2: it would be useful to explicitly state what is meant by "raw data": i.e. to explicitly state whether the data shown here include no imputation and no adjustment for imbalance in baseline utility

Response: A footnote is now included to state the meaning of ‘raw data’.

Reviewer Comment 6:

The study excludes patients who died, which could introduce bias (particularly as three quarters of the deaths were in the control group). The paper should justify excluding such patients and discuss what impact this could have had on conclusions.

Response: This point is now included in the discussion section.

Reviewer Comment 7:

Table 2, lines 39-45: it would be useful to state whether these are at follow-up (rather than baseline), for consistency with EQ-5D and to avoid confusion.

Response: We’ve deleted the ‘at follow up’ for the EQ5D to ensure consistency throughout the table.

Reviewer Comment 8:

Table 3: footnotes should be added to make it clear which of the values shown in results include imputed values and which include adjustment for which covariates. For example, do all results shown in this table include imputed values, while all differences and all of the results below the heading of "COST EFFECTIVENESS ANALYSIS" include adjustment for baseline CRQ or EQ5D score, baseline, healthcare cost and clustering? It would also be very useful to state that the cost-effectiveness analysis results are based only on the healthcare costs (not the patient costs).

Response: These suggestions have not been incorporated as additional footnotes to the table.

Reviewer Comment 9:

Table 2 and Appendix Table 3: it would be useful to expand these table headers or add a footnote to make it 100% clear what data are being presented: for example, in Appendix table 3, it would be much clearer to say "raw data estimates at baseline (including costs in the 26 weeks leading up to randomisation)" rather than "raw data estimates at baseline (26 weeks pre-randomisation)", since the latter incorrectly implies that patients were surveyed 26 weeks before they were randomised. Similarly, in Table 2, it would be clearer to have a footnote that explicitly stated that resource use

captures the period from randomisation to week 22.

Response: Both table heading have not been updated.

Reviewer Comment 10:

1. The authors have not amended the text in response to many of the reviewers' comments: e.g. R1 points 3, 37, 38 & R2 points 3, 7,10. In these cases, the authors simply replied in the response note, but did not add additional detail into the paper.

R1 Point 3: Missing Data and Imputation

Response: These issues have been explicitly addressed.

R1 Point 37: [Explain] why the single study was a sufficient source of clinical effectiveness data

RESPONSE: This is now addressed in the discussion section.

R1 Point 38: Discussion of generalisability and equity

RESPONSE: These are now addressed in the discussion section.

R2 Point 3: Intervention – Usual Care

RESPONSE: There is now a fuller description of usual care.

R2 Point 7: CRQ

RESPONSE: We believe we addressed this point in the previous revision and provide sufficient information on the nature of this variable.

R2 Point 10: Chart Searches

RESPONSE: We now clarify in the paper that these did include electronic chart searches.

Reviewer Comment 11:

2. The authors comment that "For example, the shape of the CEAC for CRQ would also likely be different if additional points between €0 and €5,000 were evaluated." Although I recognise that any choice of ceiling ratio range (particularly that for measures other than QALYs) is arbitrary, I feel that the CEACs shown for CRQ are misleading because they do not have any points in the region of the point estimate (€850/1 point change in CRQ). I feel that the CEACs should therefore be redrawn with more points between €0 and €5000.

Response: We acknowledge this point and we now include a second set of curves which present the probability of the intervention being cost effective for a range of thresholds from €0 to €5000. This is now referred to in the Discussion. We don't present a new CEAC as it is inconsistent. WE hope the presentation of the results in the Discussion will suffice.
Reviewer Comment 12:

3. Although the revised manuscript follows all of the points on the EVEREST checklist, the authors didn't revise their paper to address two points on the CHEERS checklist:
[Explain] why the single study was a sufficient source of clinical effectiveness data
Discussion of generalisability and equity

Response: In the Discussion section we now explicitly address the points of the single study design, generalizability and equity.